CS 281A/Stat 241A Homework Assignment 5 (due November 13)

1. **HMM with mixture model emissions.**

   A common modification of the HMM involves using mixture models for the emission probabilities $p(y_t|q_t)$. For concreteness, let's assume that the y_t are real-valued vectors, and thus our model involves a mixture of Gaussians for each value of the state.

   (a) Draw the graphical model for this modified HMM, identifying clearly the additional latent variables that are needed.

   (b) Write the expected complete log likelihood for the model and identify the expectations that you need to compute in the E step.

   (c) Outline an algorithm for computing the E step, relating it to the standard alpha and beta recursions.

   (d) Write down the equations that implement the M step.

2. **Conjugate duality.**

   Let $f(x)$ be a convex function. It is a fact of convex analysis that to each convex $f(x)$ there corresponds a *conjugate function* $f^*(x)$ that has the following dual relationship to $f(x)$:

$$\begin{aligned} f(x) &= \sup_{\mu}(\mu x - f^*(\mu)) \\ f^*(\mu) &= \sup_{x}(\mu x - f(x)). \end{aligned}$$

   (a) Derive the conjugate function for $f(x) = -\log(x)$.

   (b) Show that the negative of the logarithm of the logistic function is convex and derive its conjugate function.

   (c) Prove Jensen's inequality using conjugate duality.

3. **EM algorithm for Hidden Markov Models.**

   (a) Implement the EM algorithm for HMM's with Gaussian emission probabilities $p(y_t|q_t)$, where y_t is a two–dimensional real vector. Restrict the covariance matrices to be isotropic: $\Sigma = \sigma^2 I$.

   (b) Fit a HMM with 4 states to the two–dimensional data in *hmm-gauss.dat* and evaluate the log likelihood on the training and test data (*hmm-test.dat*). Plot the data together with the means of the component densities.

   (c) Fit a Gaussian mixture model with 4 states to the same data (again with isotropic covariance matrices $\sigma^2 I$). Compare the performance with that of the HMM.

4. **Factor analysis and principal component analysis**.

We have supplied you with two 2-dimensional data sets that illustrate subspace methods: *pca1.dat* and *pca2.dat*. To generate the data, we first chose a line through the origin and chose random samples from a univariate standard Gaussian distribution along that line. We then "corrupted" these data in two different ways: (1) by using an additive two-dimensional Gaussian with equal covariances in the y_1 and y_2 directions (*pca1.dat*), and (2) by using an additive two-dimensional Gaussian with greater covariance in the y_2 than in the y_1 direction (*pca2.dat*). You are to compare the factor analysis and the principal component fits to these two data sets and comment on what changes and what stays the same.

(a) Write a Matlab or SPlus implementation of PCA: For each data set you should compute the sample covariance matrix, determine the principal eigenvector, and project the data onto the corresponding subspace.

(b) Write a Matlab or SPlus implementation of factor analysis using the EM algorithm discussed in class. Once you've determined the parameters, for each data point you can compute the posterior probability $p(x|y)$; this is the factor analysis equivalent of projecting onto the principal subspace.

(c) Compute the fits for both data sets and plot the resulting projections. What changes and what stays the same?

(d) Factor analysis and PCA differ fundamentally in the way they treat variance. Comment on this point in the light of the above experiments.